1-2006

Intraday Stock Prices, Volume, and Duration: A Nonparametric Conditional Density Analysis

Anthony S. Tay
Singapore Management University, anthonytay@smu.edu.sg

Christopher Ting
Singapore Management University, christophert@smu.edu.sg

DOI: https://doi.org/10.1007/s00181-005-0004-y

Published in Empirical Economics, January 2006, Volume 30, Issue 4, pp 827-842
http://doi.org/10.1007/s00181-005-0004-y

Anthony S. Tay · Christopher Ting

Intraday stock prices, volume, and duration: a nonparametric conditional density analysis

Abstract We investigate the distribution of high-frequency price changes, conditional on trading volume and duration between trades, on four stocks traded on the New York Stock Exchange. The conditional probabilities are estimated nonparametrically using local polynomial regression methods. We find substantial skewness in the distribution of price changes, with the direction of skewness dependent on the sign of trade. We also find that the probability of larger price changes increases with volume, but only for trades that occur with longer durations. The distribution of price changes vary with duration primarily when volume is high.

1 Introduction

Time—in the form of the duration between trades—matters in the formation of stock prices. This has been demonstrated from both theoretical and empirical perspectives. Durations may be negatively related to prices because short-selling constraints prevent trading on private bad news whereas there are no similar constraints to prevent trading on private good news (Diamond and Verrecchia 1987). Durations may also be negatively related to volatility of price changes (Easley and O'Hara 1992). The connection between price change and duration has been verified empirically (Engle 2000; Grammig and Wellner 2002). Further investigations into the relationship between durations and prices have yielded interesting insights. For instance, the size and speed of price movements increase with decreasing duration (Dufour and Engle 2000), reflecting a link between duration and market liquidity. There is also an interesting dynamic relationship

A. S. Tay (✉)
School of Economics and Social Sciences, Singapore Management University,
90 Stamford Road, Singapore, 178903, Singapore
E-mail: anthonytay@smu.edu.sg

C. Ting
Lee Kong Chian School of Business, Singapore Management University, Singapore, Singapore

between price changes and duration (Russell and Engle 2004) and within duration itself (Engle and Russell 1998). Grammig and Wellner (2002) explore the interdependence of transaction intensity and volatility, and find that lagged volatility has significant negative effect on volatility in the secondary equity market following a initial public offering. The economic interpretation of the (contemporaneous) role of duration in the distribution, particularly the volatility, of price changes is the subject of a study by Renault and Werker (2004) where, using a structural model, the effect of duration on price is decomposed into a temporal effect and an informational effect. This enables them to disentangle the effects of instantaneous causality from Granger causality in the relationship between duration to prices.

As in Renault and Werker (2004) we are interested in the contemporaneous relationship between duration and price changes. We report estimates of the entire distribution of price change conditional on duration, and investigate the role of volume and trade sign on these distributions. We explore these relationships from a fully nonparametric perspective. In particular, we estimate nonparametrically the probabilities of various price changes conditional on duration, volume, lagged values of these three variables, and trade sign.

As the conditional distribution contains all probabilistic information regarding the statistical behavior of a variable given values of a set of explanatory variables, analyzing estimates of conditional distributions may reveal interesting and useful structure in the data, and provide insights that would complement studies that focus on the conditional mean or variance. The purpose for adopting a nonparametric approach to estimating the conditional distribution is to allow the data to speak for itself. There are numerous applications in the literature that highlight the value of incorporating nonparametric density estimation into an analysis. For example, Gallant et al. (1992) study the bivariate distribution of daily returns and volume conditional on lagged values of these variables. Among other findings, their analysis indicates a positive correlation between risk and return after conditioning on lagged volume. Other studies that highlight the usefulness of non-parametric methods include Engle and Gonzalez-Rivera (1991), and Gallant et al. (1991).

We estimate the conditional distributions for four stocks traded on the NYSE using intraday data from TAQ spanning a period of 1 year, from Jan 2, 2002 to Dec 30, 2002. We find substantial skewness in the distribution of price changes, with the direction of skewness dependent on trade sign. On the whole, the relationship between price changes and volume is much weaker than the relationship between price changes and duration, and shows up most clearly at long durations. When durations are long, the probability of large price changes increases with volume.

In the following section we describe the data used in this study, and explain precisely all the adjustments made to the data to get it into a form suitable for analysis. In Section 3, we describe the nonparametric conditional distribution estimation technique used in the paper, and discuss the practical issues we had to address in order to the implement the technique. The results of our study are presented in Section 4, followed by concluding comments.

2 Data and data adjustments

We estimate the conditional distributions of price changes for four NYSE stocks: IBM (International Business Machines), GE (General Electric), BA (Boeing), and

MO (the Altria Group, formerly Philip Morris.) The data are obtained from the TAQ database, and cover the period Jan 2, 2002 to Dec 30, 2002. We extract, for each stock, the time of trade of the tth transaction τ_t, from which we obtain the duration of the tth transaction $d_t = \tau_t - \tau_{t-1}$ (τ_0=34,200 s after midnight, i.e., 9:30 am), the transacted price p_t, from which we compute the price change $\Delta p_t = p_t - p_{t-1}$, and volume v_t in lots of 100 shares. Clearly, these are all discrete variables. Each trade is then signed as in Lee and Ready (1991) to indicate if the transaction is buyer (+) or seller (−) initiated, but without modifying the reported times of quotes as developments in the NYSE trading procedures no longer warrant this. There are newly proposed methods (e.g., Vergote 2005) that aim to determine the appropriate adjustments to the times of quotes. We do not apply these techniques, but as the analysis in Vergote (2005) suggests that a delay of 1 or 2 s may be appropriate in the sample period that we work with, we check if adding 2 s to the quote time-stamp affects our results.

We use data from 0930 to 1600 h, deleting all trades that occur outside of these hours. Data from 4 days with unusual market openings and closings are dropped from our sample. These are July 5, September 11, November 29, and December 24, 2002. The first of these dates is an early closing for Independence Day, the second is a late opening, in respect of memorial events commemorating the 1-year anniversary of the attacks on World Trade Center. The latter 2 days are early closings for Thanksgiving and Christmas.

There are several noteworthy characteristics of the data set in our sample period. One is that by the start of this sample period the NYSE had already completed the move to decimal pricing. One benefit of this is that the bid-ask spreads are small so that that the bid-ask bounce is less of an issue for our estimates. Perhaps the more important characteristic of this sample period is that, in general, trading is so active that for each stock in our study there are large numbers of trades with zero duration. These may be trades that occur almost simultaneously, but are recorded as having occurred at the same time because the TAQ database records time of trade to an accuracy of 1 s. Some of these trades may also reflect large trades that are broken up into smaller simultaneous trades. The exact number of such trades for the stocks we analyze are presented in Table 1. The lowest proportion of zero-duration trades is BA at 3.7%. About 5% of the IBM and MO observations have zero duration, and almost 10% of the observations for GE have zero duration. We aggregate in standard fashion all trades that occur with the same time-stamp and consider the aggregate as a single trade. The price of the first trade in the aggregate is taken as the price of the aggregate trade. Signed volume is

Table 1 Number of zero duration trades

	IBM	GE	BA	MO
Total number of observations	923,577	1,292,532	594,186	797,373
Number of zero durations	54,799	121,924	21,925	41,651
	(5.9%)	(9.4%)	(3.7%)	(5.2%)
Number of observations after aggregation	868,778	1,170,608	572,261	755,722

simply aggregated. Even after aggregation we have a large number of observations for each stock, ranging from 572,261 (BA) to 1,170,608 (GE). Finally, note that, unlike many studies that work with intraday stock prices, we do not remove the diurnal patterns that are present in durations. This is a more critical issue if the focus of the study involves the dynamics in duration data, but our focus is on the contemporaneous relationship between durations and price changes, and so we choose to examine the data with as few adjustments as possible. In addition, we do not estimate the distributions at a fine enough grid on durations for the removal of diurnal patterns to affect our results in any important way.

In Fig. 1 we show histograms of price changes, duration, and signed volume for IBM. Also displayed is the histogram for signed duration (duration multiplied by the sign of the trade). The distribution of price change appears to be very symmetric. The distribution of signed durations is symmetric, so the distribution of durations is similar for both buyer- and seller-initiated trades. The histogram for trading volume is not informative except to indicate the presence of a few outliers. However, apart from these the distribution is also fairly symmetric. A more detailed picture can be obtained from Table 2. Here we show every 10th percentile of price change, (unsigned) duration, duration for buyer- and seller-initiated trades, (unsigned) volume, and volume for buyer- and seller-initiated trades. For all stocks the distribution of duration and volume across buyer- and seller-initiated trades are very similar to the overall distribution. In fact there is almost no correlation between the sign of trade and duration. Volume is substantially larger for the more

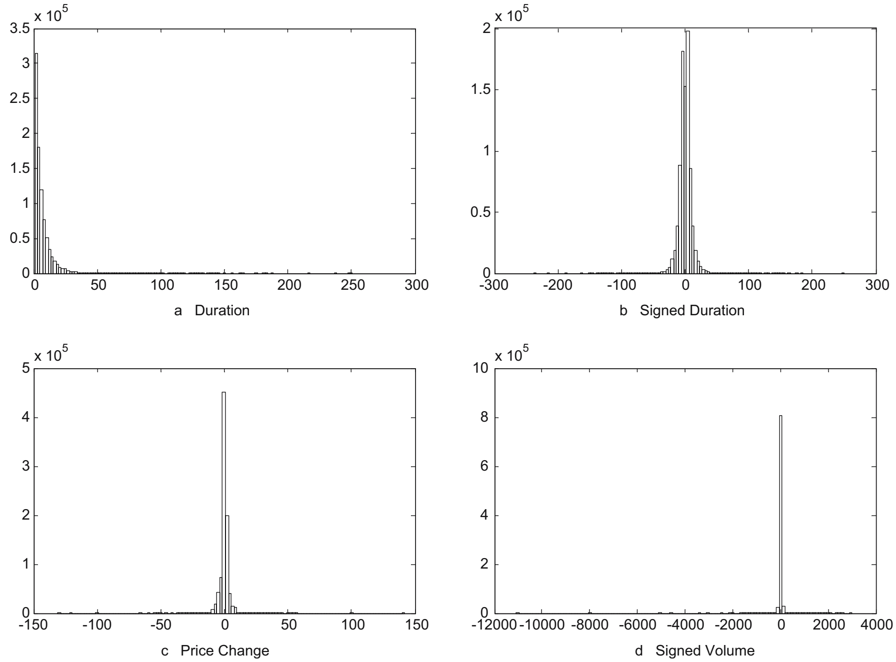


Fig. 1 Histograms of duration, price changes, and volume (*IBM*)

Table 2 Distribution of prices, duration, and volume

		Percentiles								
		10	20	30	40	50	60	70	80	90
IBM	Price change	−3	−1	0	0	0	0	1	1	3
	Duration	1	2	3	4	5	6	7	10	14
	Duration (sell)	2	2	3	4	5	6	7	10	14
	Duration (buy)	1	2	3	4	5	6	7	10	14
	Volume	1	2	4	5	7	10	14	23	48
	Volume (sell)	1	2	4	5	7	10	13	22	45
	Volume (buy)	1	2	4	5	7	10	15	24	50
GE	Price change	−1	−1	0	0	0	0	0	1	1
	Duration	1	1	2	2	3	4	5	7	11
	Duration (sell)	1	1	2	2	3	4	5	7	11
	Duration (buy)	1	1	2	2	3	4	5	7	11
	Volume	1	3	5	8	10	15	24	42	94
	Volume (sell)	1	3	5	8	10	15	24	40	90
	Volume (buy)	1	3	5	8	10	15	24	43	99
BA	Price change	−2	−1	0	0	0	0	0	1	2
	Duration	1	2	3	4	5	7	10	15	25
	Duration (sell)	1	2	3	4	5	7	10	15	25
	Duration (buy)	1	2	3	4	5	7	10	15	25
	Volume	1	2	2	3	5	7	10	13	24
	Volume (sell)	1	2	2	3	5	6	10	12	23
	Volume (buy)	1	2	2	4	5	7	10	13	25
MO	Price change	−2	−1	0	0	0	0	0	1	2
	Duration	1	2	3	3	5	6	8	11	18
	Duration (sell)	1	2	3	3	5	6	8	11	18
	Duration (buy)	1	2	3	3	5	6	8	11	17
	Volume	1	2	3	5	6	9	12	20	42
	Volume (sell)	1	2	3	5	6	9	12	20	41
	Volume (buy)	1	2	3	5	6	9	12	20	43

frequently traded of the four stocks. The distribution of price changes is mostly symmetric. Table 2 also provides a useful reference for interpreting our conditional distribution plots, as our conditional probabilities will be estimated and plotted at various *percentile* values of the conditioning information set.

3 Nonparametric estimation of conditional distribution functions

Let $\{\mathbf{X}_t, Y_t\}_{t=1}^{T}$ be observations from a strictly stationary process where Y_t is a scalar and $\mathbf{X}_t = (x_{1t}, x_{2t}, \ldots x_{Mt})$. In our first application, $Y_t = \Delta p_t$ and $\mathbf{X}_t = (d_t, sv_t)$. One nonparametric approach to estimating the conditional distribution function

$$F(y|\mathbf{x}) = \Pr(Y_t \leq y|\mathbf{X}_t = \mathbf{x})$$

is to make use of the fact that if $Z_t = I(Y_t \leq y)$, where $I(.)$ is the indicator function, then $E\,[Z_t \mid \mathbf{X}_t = \mathbf{x}] = F(y \mid \mathbf{x})$. The particular technique we use is the Adjusted Nadaraya–Watson estimator (Hall et al. 1999; Hall and Presnell 1999) which we state in its multivariate ($M \geq 1$) form.

3.1 The Adjusted Nadaraya–Watson estimator

The Adjusted Nadaraya–Watson estimator of $F(y|\mathbf{x})$ is given by

$$\tilde{F}(y|\mathbf{x}) = \frac{\sum_{t=1}^{T} Z_t w_t K_{\mathbf{H}}(\mathbf{X}_t - \mathbf{x})}{\sum_{t=1}^{T} w_t K_{\mathbf{H}}(\mathbf{X}_t - \mathbf{x})} \tag{1}$$

where $\{w_t\}_{t=1}^{T} = \arg\max \prod_{t=1}^{T} w_t$, with $\{w_t\}_{t=1}^{T}$ satisfying the conditions (1) $w_t \geq 0$ for all t, (2) $\sum_{t=1}^{T} w_t = 1$, and (3) $\sum_{t=1}^{T} w_t (X_{mt} - x_m)\, K_{\mathrm{H}}(\mathbf{X}_t - \mathbf{x}) = 0$ for all $m = 1,\ldots,M$, and $K_{\mathrm{H}}(.)$ is a multivariate kernel with bandwidth matrix $\mathbf{H}$.

Although the Adjusted Nadaraya–Watson estimator is based on the biased bootstrap idea of Hall and Presnell (1999), it is useful, as noted in Hall et al. (1999), to view the estimator as the local linear estimator of $F(y|\mathbf{x})$ with weights $K_{\mathbf{H}}(\mathbf{X}_t - \mathbf{x})$ replaced by $w_t K_{\mathbf{H}}(\mathbf{X}_t - \mathbf{x})$, i.e., $\tilde{F}(y|\mathbf{x}) = \hat{a}$ where $\hat{a}$ is obtained from the solution of

$$\max_{a,\,\mathbf{b}} \sum_{t=1}^{T} (Z_t - a - (\mathbf{X}_t - \mathbf{x})\mathbf{b})^2 w_t K_{\mathbf{H}}(\mathbf{X}_t - \mathbf{x})$$

(see Fan and Gijbels 1996, for an authoritative introduction to local linear and local polynomial regression methods). It is easy to see from the first-order condition

$$\frac{\partial}{\partial a} \sum_{t=1}^{T} (Z_t - a - (\mathbf{X}_t - \mathbf{x})\mathbf{b})^2 w_t K_{\mathbf{H}}(\mathbf{X}_t - \mathbf{x}) = 0$$

that $\hat{a}$ reduces to $\tilde{F}(y|\mathbf{x})$ under condition (3).

Using the unmodified version of the local linear approach ($w_t = 1$) may result in estimates of conditional distributions that are not monotonic in y, or that do not lie always between 0 and 1. The Adjusted Nadaraya–Watson estimates, on the other hand, always lie between 0 and 1, is monotonic in y, and yet share the superior bias properties as estimates from local linear methods (Hall et al. 1999), and also automatic adaptation to estimation at the boundaries (see e.g., Fan and Gijbels 1996). There is no requirement for the conditional distribution to be continuous in y. Another justification for using the adjusted Nadaraya–Watson estimator is provided by Cai (2002) who establish asymptotic normality and weak consistency of the estimator for time series data under conditions more general than in Hall et al. (1999).

3.2 Practical issues

Implementation of the estimator $\tilde{F}(y|\mathbf{x})$ requires a number of practical issues to be addressed. In particular, $\{p_t\}$ has to be computed, and $K_{\mathbf{H}}(.) = |\mathbf{H}|^{-1} K(\mathbf{H}^{-1}\mathbf{x})$ has to be chosen. The choice of K in smoothing problems is usually not crucial (see e.g., Wand and Jones 1995) but the choice of $\mathbf{H}$ is important. For K we use the standard M-variate normal distribution

$$K(\mathbf{x}) = (2\pi)^{-1/2} \exp\left(-\|\mathbf{x}\|^2 \big/ 2\right).$$

We take $\mathbf{H}$=$h\mathbf{I}_M$ where $\mathbf{I}_M$ is the M-dimensional identity matrix. Other, possibly non-symmetric kernels may be useful here (for instance, the gamma kernels in Chen 2000), although we stay with the Gaussian kernel as the theoretical properties for this specific method has been established for symmetric kernels only. As we are effectively using only one bandwidth for multiple regressors, our regressors are always scaled to a common variance before the estimator is implemented.

To obtain the optimal value of h, we adapt the bootstrap bandwidth selection method suggested by Hall et al. (1999). This approach exploits the fact that, as we are estimating distribution functions, there is limited scope for highly complicated behavior. First, a simple parametric model is fitted to the data and used to obtain an estimate $\hat{F}(y|\mathbf{x})$. We use

$$Y_t = a_0 + a_1 X_{1t} + \ldots + a_M X_{Mt} + a_{M+1} X_{1t}^2 + \ldots + a_{2M} X_{Mt}^2 + \varepsilon_t \qquad (2)$$

and assume that ε_t is heteroskedastic, depending on the square of lagged price changes. We then simulate from this model to obtain a bootstrap sample $\{Y_1^*, Y_2^*,\ldots,Y_T^*\}$ using the actual observations $\{\mathbf{x}_1,\mathbf{x}_2,\ldots\mathbf{x}_T\}$. For each bootstrap sample (and for a given value of h), we compute a bootstrap estimate $\tilde{F}_h^*(y|\mathbf{x})$. We choose h to minimize

$$\sum \left|\tilde{F}_h^*(y|\mathbf{x}) - \hat{F}(y|\mathbf{x})\right|$$

where the summation is over all bootstrap replications and over values of y for any given $\mathbf{x}$. We checked the sensitivity of our estimates to the choice of the bandwidth, and we note that we obtained very similar results over a wide range of bandwidth values.

Computation of the weights w_t is carried out using the Lagrange Multiplier method. The Lagrangian is

$$L = \sum_{t=1}^{T} \log(w_t) - \lambda_0 \left(\sum_{t=1}^{T} w_t - 1\right) - \sum_{m=1}^{M} \lambda_m \sum_{t=1}^{T} w_t (X_{m,t} - x_m) K_{\mathbf{H}}(\mathbf{X}_t - \mathbf{x}),$$

which gives the first-order conditions

$$\frac{1}{w_t} - \lambda_0 - \sum_{m=1}^{M} \lambda_m (X_{m,t} - x_m) K_{\mathbf{H}}(\mathbf{X}_t - \mathbf{x}) = 0,\ \forall m = 1, \ldots, M \qquad (3)$$

together with the restrictions $\sum_{t=1}^{T} w_t=1$ and $\sum_{t=1}^{T} w_t(X_{mt}-x_m)\ K_{\mathbf{H}}(\mathbf{X}_t-\mathbf{x})=0$ $\forall m=1,...,M$. Solving these equations, we get $\lambda_0=T$ with $\{\lambda_m\}_{m=1}^{M}$ satisfying the equations

$$\sum_{t=1}^{T} \frac{(X_{mt}-x_m)K_{\mathbf{H}}(\mathbf{X}_t-\mathbf{x})}{T+\sum_{m=1}^{M}\lambda_m(X_{mt}-x_m)K_{\mathbf{H}}(\mathbf{X}_t-\mathbf{x})}=0, \quad \forall m=1,\ldots,M \tag{4}$$

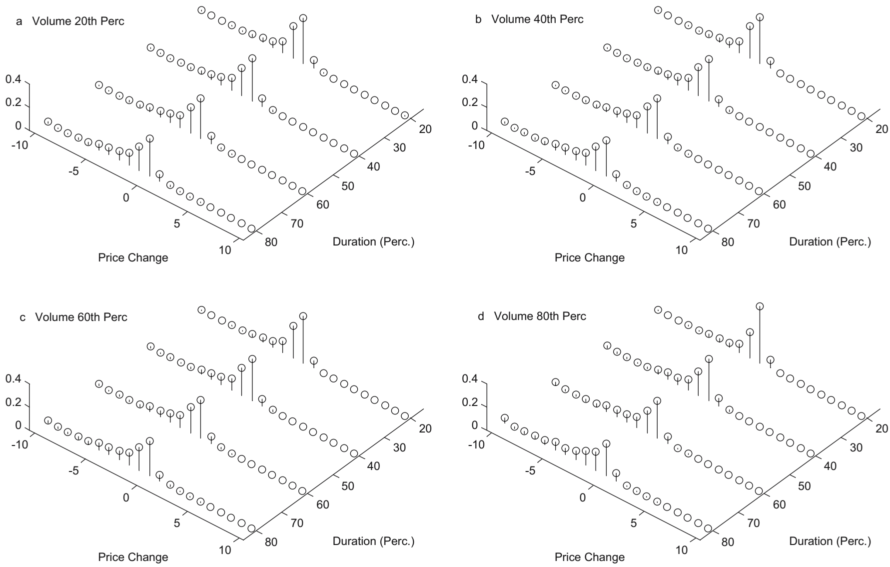
We obtain $\{\lambda_m\}_{m=1}^{M}$ by solving Eq. (4) numerically (we use the MATLAB function *fsolve* to do this.) The weights $\{w_t\}_{t=1}^{T}$ are then computed from Eq. (3).

4 Empirical results

For each of the four stocks in our sample, we estimate two sets of the conditional distribution $F(\Delta p_t|d_t,v_t)$, one set for seller-initiated trades, and one for buyer-initiated trades. For each set, the distribution is estimated at the 20th, 40th, 60th, and 80th percentiles of d_t and v_t. That is, for each stock, we estimate 16 conditional distributions. It is not worth the computational burden to estimate the conditional distributions at a finer grid: it would be difficult to present that much information clearly and simply, and the chosen values of the conditioning variables are sufficient for highlighting important empirical regularities. For each pair (d_t,v_t) we estimate the conditional cumulative distribution at values of Δp_t from −10 to 9. We report the conditional probabilities $\Pr(\Delta p_t\leq-10|d_t,v_t)$, $\Pr(\Delta p_t=i|d_t,v_t)$ for $i=-9, -8,...,8, 9$, and $\Pr(\Delta p_t\geq 10|d_t,v_t)$. These conditional probabilities are obtained by taking the difference of the estimated cumulative distribution between adjacent points of the grid over Δp_t. It is possible to go further and obtain estimates of the bivariate conditional distribution $\Pr(\Delta p_t,d_t|v_t)$ by first estimating $\Pr(d_t|v_t)$, and getting the distribution $\Pr(\Delta p_t,d_t|v_t)$ by multiplying the estimates of $\Pr(d_t|v_t)$ with the estimates of $\Pr(\Delta p_t|d_t,v_t)$ obtained previously. Our estimates of $\Pr(d_t|v_t)$ show that more trades occur at short durations than at long durations, regardless of volume, and as a result, much of the interesting structure in the conditional probabilities $\Pr(\Delta p_t|d_t,v_t)$ do not show up well in plots of the bivariate distributions. We therefore discuss only our estimates of $\Pr(\Delta p_t|d_t,v_t)$.

The results for IBM are presented in Fig. 2(a) and (b). In Fig. 2(a), we show the distributions of Δp_t conditional on d_t and v_t for seller-initiated trades, organized into four panels. In each panel we have the conditional distribution at four percentile levels of duration (20th, 40th, 60th, and 80th percentiles—see Table 2 for actual values.) The four panels correspond to four different levels of volume, with volume at the 20th, 40th, 60th, and 80th percentile levels starting from the top left panel and proceeding clockwise.) In Fig. 2(b) we have the same plots for buyer-initiated trades. The most obvious pattern is that the distributions in Fig. 2(a) are negatively skewed, whereas the distributions for buyer-initiated trades in Fig. 2(b) are positively skewed. This skewness is also clear in the estimated probabilities provided later in Table 3. As security prices tend to move in the direction of trade sign, the skewness observed is not surprising. Buyer-initiated trades tend to move the security price higher, especially when these trades exhaust the prevailing depth in the limit-order book and NYSE specialists revise the quotes higher to reflect the higher demand for the stock.

a



b

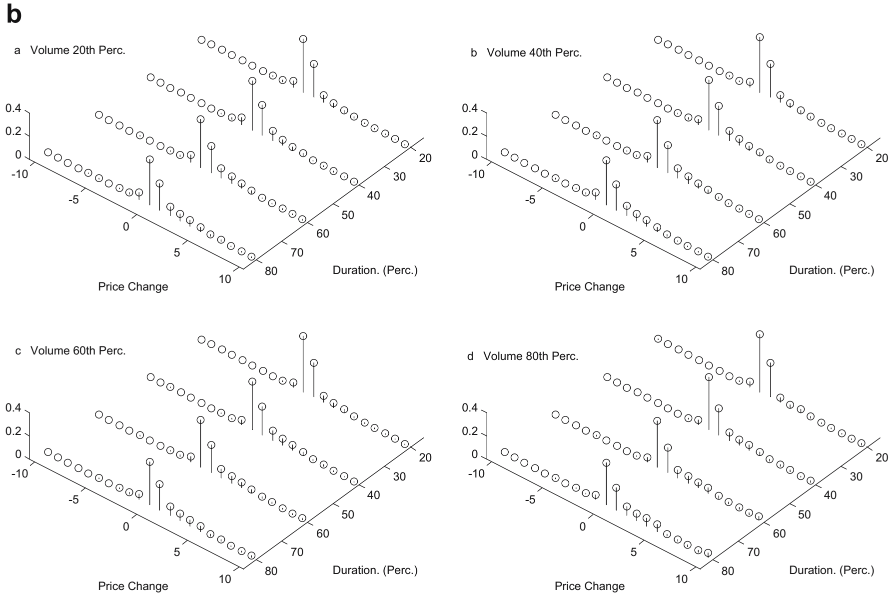


Fig. 2 **a** Estimated conditional probabilities (*IBM* Seller-initiated trades). **b** Estimated conditional probabilities (*IBM* Buyer-initiated trades)

Table 3 Estimated probabilities of price change conditional on volume and duration

	Volume (percentile)	Duration (percentile)	$\Delta p_t \leq -2$				$-1 \leq \Delta p_t \leq 1$				$\Delta p_t \geq -2$			
			IBM	GE	BA	MO	IBM	GE	BA	MO	IBM	GE	BA	MO
(a) Seller-initiated trades	20th	20th	0.243	0.119	0.139	0.148	0.742	0.875	0.848	0.839	0.016	0.006	0.014	0.013
		40th	0.298	0.142	0.203	0.184	0.677	0.850	0.775	0.799	0.025	0.008	0.022	0.017
		60th	0.327	0.157	0.222	0.199	0.645	0.834	0.751	0.784	0.027	0.008	0.027	0.017
		80th	0.374	0.163	0.226	0.210	0.601	0.831	0.738	0.774	0.025	0.006	0.036	0.016
	40th	20th	0.246	0.125	0.141	0.158	0.742	0.869	0.846	0.832	0.012	0.006	0.013	0.010
		40th	0.307	0.146	0.205	0.192	0.673	0.846	0.775	0.796	0.020	0.008	0.021	0.012
		60th	0.341	0.163	0.227	0.215	0.636	0.829	0.748	0.772	0.024	0.008	0.025	0.013
		80th	0.390	0.168	0.253	0.224	0.583	0.825	0.717	0.763	0.027	0.007	0.030	0.014
	60th	20th	0.241	0.128	0.142	0.163	0.749	0.866	0.847	0.829	0.011	0.007	0.012	0.009
		40th	0.320	0.153	0.212	0.200	0.661	0.839	0.769	0.790	0.018	0.008	0.019	0.011
		60th	0.366	0.174	0.241	0.231	0.612	0.818	0.737	0.757	0.022	0.008	0.023	0.012
		80th	0.411	0.181	0.269	0.245	0.564	0.812	0.704	0.746	0.026	0.007	0.027	0.010
	80th	20th	0.202	0.120	0.133	0.161	0.789	0.874	0.858	0.831	0.009	0.005	0.009	0.008
		40th	0.341	0.153	0.223	0.202	0.640	0.840	0.759	0.788	0.018	0.007	0.018	0.010
		60th	0.397	0.202	0.267	0.264	0.582	0.791	0.712	0.723	0.021	0.006	0.021	0.013
		80th	0.472	0.215	0.313	0.269	0.509	0.777	0.663	0.720	0.018	0.008	0.024	0.011

Table 3 (continued)

	Volume (percentile)	Duration (percentile)	$\Delta p_t \leq -2$				$-1 \leq \Delta p_t \leq 1$				$\Delta p_t \geq -2$			
			IBM	GE	BA	MO	IBM	GE	BA	MO	IBM	GE	BA	MO
(b) Buyer-initiated trades	20th	20th	0.021	0.006	0.016	0.015	0.804	0.895	0.868	0.843	0.176	0.099	0.116	0.142
		40th	0.030	0.008	0.023	0.019	0.743	0.868	0.802	0.809	0.227	0.124	0.175	0.171
		60th	0.033	0.007	0.022	0.015	0.717	0.854	0.799	0.834	0.250	0.140	0.179	0.151
		80th	0.037	0.007	0.031	0.013	0.695	0.855	0.783	0.822	0.269	0.138	0.187	0.165
	40th	20th	0.015	0.006	0.014	0.012	0.813	0.891	0.857	0.841	0.171	0.103	0.129	0.148
		40th	0.026	0.008	0.021	0.015	0.729	0.861	0.786	0.806	0.245	0.131	0.193	0.179
		60th	0.030	0.008	0.022	0.012	0.700	0.846	0.784	0.819	0.269	0.146	0.194	0.169
		80th	0.035	0.007	0.027	0.012	0.676	0.852	0.776	0.821	0.288	0.141	0.198	0.168
	60th	20th	0.013	0.006	0.011	0.010	0.816	0.886	0.890	0.837	0.171	0.108	0.100	0.154
		40th	0.024	0.008	0.019	0.012	0.701	0.848	0.776	0.793	0.275	0.145	0.206	0.195
		60th	0.027	0.008	0.021	0.011	0.694	0.831	0.783	0.807	0.279	0.161	0.197	0.182
		80th	0.033	0.007	0.025	0.010	0.650	0.838	0.766	0.798	0.317	0.155	0.210	0.192
	80th	20th	0.015	0.005	0.012	0.007	0.844	0.888	0.866	0.844	0.142	0.107	0.123	0.149
		40th	0.019	0.009	0.017	0.010	0.719	0.866	0.729	0.767	0.261	0.125	0.254	0.223
		60th	0.024	0.007	0.020	0.011	0.657	0.795	0.765	0.789	0.319	0.199	0.215	0.200
		80th	0.030	0.008	0.021	0.007	0.602	0.804	0.745	0.743	0.368	0.188	0.234	0.250

More interesting is the observation, for both buyer- and seller-initiated trades, that duration has a much stronger influence on the distribution of price change at high volumes than at low volumes. Looking at panel (a) of Fig. 2(a) and (b) we see a slight change in the distribution of price changes between the four duration levels at low volume. Panel (d) of both figures on the other hand show that at high volume, the probability of larger price changes increases substantially with duration. The influence of volume on price change also depends on duration. At short durations (i.e., when trading is very active), volume does not have very much influence on the distribution of price changes; moving from panel (a) to panel (d), in both diagrams, the distributions at the 20th percentile level of duration are very similar. At long durations, volume matters. Again moving from panel (a) to panel (d), the distribution of price change for duration at the 80th percentile shows that the probability of larger prices changes (negative for seller-initiated trades, positive for buyer-initiated ones) increases as volume increases. Duration between trades influences the distribution of price changes, but the degree of influence appears to depend on the volume.

This result suggests that at the intraday frequency, volume may have some influence on the distribution of price changes, even after controlling for duration. This contrasts with the result in Jones et al. (1994) that, at the daily frequency, the relationship between volume and the volatility of stock returns actually reflects the relationship between volatility and the number of transactions (thus, average duration). The result is consistent with research on high-frequency stock prices (e.g., Easley et al. 1997) and suggests the importance of the interaction between volume and duration.

We repeat this exercise with the other three stocks in our sample. The figures are qualitatively the same as Fig. 2, so we do not display them. Instead, the estimated probabilities $\Pr(\Delta p_t \leq -2|d_t,v_t)$, $\Pr(-1 \leq \Delta p_t \leq 1|d_t,v_t)$, and $\Pr(\Delta p_t \geq -2|d_t,v_t)$ are presented in Table 3. The top half of the table shows estimates for seller-initiated trades, while the lower half shows estimates for buyer-initiated trades.

We see that for IBM at the 20th percentile of duration, $\Pr(\Delta p_t, \leq -2|d_t,v_t)$ remains around 0.2 for seller-initiated trades, and $\Pr(\Delta p_t \geq -2|d_t,v_t)$ lies mostly around 0.17 for buyer-initiated trades, as volume increases. At the 80th percentile level of duration, both probabilities increase substantially as volume increases. Focusing on the 20th percentile level of volume, $\Pr(\Delta p_t \leq -2|d_t,v_t)$ increases from 0.243 to 0.374 as duration increases from the 20th percentile level to the 80th. At the 80th percentile level of volume, $\Pr(\Delta p_t \leq -2|d_t,v_t)$ increases from 0.202 to 0.472 as duration increases from the 20th to 80th percentile levels.

Looking over the estimates for other three stocks, we see that similar comments can be made, with differences only in degree. These patterns appear to be strongest in the case of IBM, and weakest for GE and BA. In addition to the estimates for the full sample period, we also compute the estimates for two subsamples (Jan to June, and July to December) to check if our results are robust across different sample periods. The estimates for these two sample periods are very similar to the estimates for the full sample. We also checked if adding a 2 s delay to the reported quote time when signing trades affects our results. Again, the estimates in this case are very similar to what is reported here.

We have also computed 95% confidence intervals around the probability estimates in Table 3 using formulas that exploit the local linear nature of the estimates (see e.g. Fan and Yao 2003, Section 6.3.4). We do not display the intervals, because it suffices to note that they are all very narrow-the boundary of the intervals differ from the estimated probabilities only in the third or fourth decimal places-and therefore indicate that the difference between the estimated probabilities conditional on long versus short durations are statistically significant. For instance, the 95% confidence interval for Prob($\Delta p_t \leq -2$) for IBM (volume at 80th perc., duration 80th perc.) is (0.471, 0.474) whereas the corresponding interval for IBM (volume at 80th perc., duration 20th perc.) is (0.201, 0.202).

It seems therefore that there are substantial interaction effects between duration and volume, and the sign of trade. Any parametric analysis of the relationship between duration, volume, and prices should take these interaction effects into account. In addition, the high degree of skewness in the distributions indicates that care should be taken in interpreting results concerning the volatility of price changes. We interpret our results for seller-initiated trades as evidence in support of the Diamond and Verrecchia (1987) analysis, where we expect larger probability of price falls with higher levels of duration.

As a further robustness check, we re-estimate the conditional probabilities, including lagged values of duration, price changes, and volume, in addition to contemporaneous duration and volume. We continue to report the estimated probabilities at the 20th, 40th, 60th, and 80th percentile levels of duration and volume, but only at the median values of the lagged variables. Estimates of Pr($\Delta p_t \leq -2$) for seller-initiated trades, and Pr($\Delta p_t \geq -2$) for buyer-initiated trades are listed in Table 4. We note that by conditioning on the median value of lagged price change (which is zero), the estimated probabilities of price changes of −1 to 1 ticks increase substantially; this is because trades at the same price tend to be following by more trades at the same price. Nonetheless, it is still the case that among the probabilities Pr($\Delta p_t \leq -2$) for seller-initiated trades, and Pr($\Delta p_t \geq -2$) for buyer-initiated trades, the largest estimates occur when duration and volume are both at the 60th or 80th percentiles.

The focus of this paper is on the contemporaneous relationships between duration, volume, and price changes. Nonetheless, it is of interest to extend the analysis to the dynamic interrelationships between these variables (beyond the robustness check reported in the previous paragraph) as much of the interest in these variables arise from the effect of information arrival, and changes over time of durations are indicative of this. For instance, it is of interest to explore and compare the densities of price changes when trading is becoming more frequent, with the densities when trading is becoming less frequent.

In Fig. 3, we show the estimated probabilities for GE in these two situations. The densities for the other three stocks are similar, and are omitted. The top row shows the estimated price densities for seller-initiated trades, whereas the bottom row shows the estimates for buyer initiated trades. The left column shows the estimates for the situation when trades are getting more frequent (current duration at the 20th percentile, lagged duration at the 80th percentile). Again the skewness in the distributions is clear, although in the case of GE, the modal price change is now −1 and 1 for seller and buyer initiated trades, respectively. In the other three stocks, the probability of a price change also increases relative to the probability of no price change.

In the right column, we show the estimated probabilities when current duration is long relative to lagged duration (current duration at the 80th percentile, lagged duration at the 20th percentile). Here the modal probability is clearly at zero price

Table 4 Estimated probabilities of price change conditional on volume and duration and lagged information

	Volume (Perc.)	Duration (Perc.)	$\Delta p_t \leq -2$					Volume (Perc.)	Duration (Perc.)	$\Delta p_t \geq -2$			
			IBM	GE	BA	MO				IBM	GE	BA	MO
(a) Seller-initiated trades	20th	20th	0.070	0.026	0.074	0.096	(b) Buyer-initiated trades	20th	20th	0.073	0.025	0.044	0.048
		40th	0.026	0.046	0.141	0.067			40th	0.129	0.036	0.167	0.067
		60th	0.245	0.053	0.116	0.059			60th	0.084	0.029	0.070	0.040
		80th	0.156	0.107	0.111	0.033			80th	0.119	0.023	0.092	0.066
	40th	20th	0.175	0.034	0.066	0.135		40th	20th	0.097	0.031	0.055	0.051
		40th	0.118	0.052	0.117	0.087			40th	0.120	0.044	0.097	0.074
		60th	0.151	0.118	0.113	0.098			60th	0.138	0.041	0.114	0.079
		80th	0.211	0.046	0.109	0.059			80th	0.172	0.040	0.102	0.065
	60th	20th	0.113	0.054	0.065	0.064		60th	20th	0.088	0.036	0.054	0.057
		40th	0.202	0.067	0.136	0.095			40th	0.148	0.048	0.115	0.086
		60th	0.427	0.077	0.183	0.104			60th	0.199	0.055	0.095	0.081
		80th	0.324	0.087	0.189	0.082			80th	0.195	0.057	0.089	0.061
	80th	20th	0.082	0.038	0.084	0.086		80th	20th	0.076	0.022	0.156	0.085
		40th	0.235	0.062	0.176	0.146			40th	0.243	0.049	0.126	0.112
		60th	0.254	0.362	0.192	0.108			60th	0.447	0.136	0.260	0.132
		80th	0.215	0.088	0.130	0.108			80th	0.272	0.285	0.117	0.086

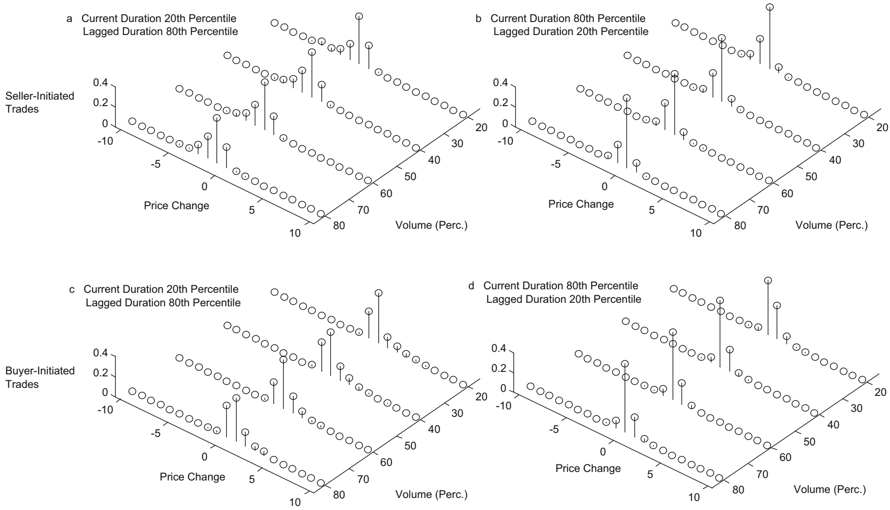


Fig. 3 Estimated conditional probabilities (*GE* with lagged duration)

change. The distributions here are also much tighter than the corresponding distributions in Fig. 2. This suggests that our results reported earlier are more appropriate at median values of the lagged variables, or when current durations are short relative to lagged durations. These results are merely indicative; a more extensive study of the dynamic relationships in the framework used here is left for future research.

5 Concluding comments

We investigate the distribution, conditional on trading volume, duration between trades, and the sign of trades of high-frequency price changes on four stocks traded on the New York Stock Exchange. The conditional probabilities are estimated nonparametrically using local polynomial regression methods. We find substantial skewness in the distribution of price changes, with the direction of skewness dependent on trade sign. We also find that the probability of larger price changes increases with volume, but only for trades that occur with longer durations. Durations affect prices, with a stronger effect when volume is high.

The evidence suggests substantial interaction effects between duration and volume with respect to their effect on prices; parametric analyses of the relationship between duration, volume, and prices should take these into account, such as in Tay et al. (2004). The high degree of skewness in the distributions indicate that it is also important to distinguish between seller- and buyer-initiated trades, and that care should be taken in interpreting results concerning the volatility of price changes. The results are consistent with theoretical research. Our findings for seller-initiated trades provide direct evidence in support of the Diamond and Verrecchia (1987) analysis, where larger probability of price falls is associated with a higher level of duration.

Acknowledgements Tay gratefully acknowledges research support from the Singapore Management University's Wharton-SMU Research Centre. We thank participants at the 2005 North American Meeting of the Econometric Society for helpful questions and comments. All errors remain our own. Zhang Feng provided able research assistance.

References

Cai Z (2002) Regression quantiles for time series. Econom Theory 18:169–192
Chen SX (2000) Probability density function estimation using gamma kernels. Ann Inst Stat Math 52:471–480
Diamond DW, Verrecchia RE (1987) Constraints on short-selling and asset price adjustment to private information. J Financ Econ 18:277–311
Dufour A, Engle RF (2000) Time and the price impact of a trade. J Finance 55(6):2467–2498
Easley D, O'Hara M (1992) Time and the process of security price adjustment. J Finance 47 (2):577–605
Easley D, Kiefer NM, O'Hara M (1997) The information content of the trading process. J Empir Finance 4:159–186
Engle RF (2000) The econometrics of ultra-high frequency data. Econometrica 68(1):1–22
Engle RF, Gonzalez-Rivera G (1991) Semiparametric ARCH models. J Bus Econ Stat 9:345–360
Engle RF, Russell JR (1998) Autoregressive conditional duration: a new model for irregularly spaced transaction data. Econometrica 66(5):1127–1162
Fan J, Gijbels I (1996) Local polynomial modelling and its application, monographs on statistics and applied probability 66, Chapman & Hall/CRC
Fan J, Yao Q (2003), Nonlinear time series: nonparametric and parametric methods, Springer
Gallant AR, Hsieh DA, Tauchen G (1991) On fitting a recalcitrant series: the Pound/Dollar exchange rate, 1974–83. In: Barnett WA, Powell J, Tauchen G (eds) Nonparametric and semiparametric methods in econometrics and statistics: proceedings of the fifth international symposium in economic theory and econometrics, Cambridge University Press
Gallant AR, Rossi PE, Tauchen G (1992) Stock prices and volume. Rev Financ Stud 5(2):199–242
Grammig J, Wellner M (2002) Modeling the interdependence of volatility and inter-transaction duration processes. J Econom 106:369–400
Hall P, Presnell B, Intentionally Biased Bootstrap Methods. J R Stat Soc B 61:143–158
Hall P, Wolff RCL, Yao Q (1999) Methods for estimating a conditional distribution function. J Am Stat Assoc 94(445)154–163
Jones CM, Kaul G, Lipson ML (1994) Transactions, volume, and volatility. Rev Financ Stud 7 (4):631–651
Lee CMC, Ready MJ (1991) Inferring trade direction from intraday data. J Finance 46(2)733–746
Renault E, Werker BJM (2004) Stochastic Volatility Models with Transaction Time Risk, manuscript, Universite de Montreal and Tilburg University
Russell JR, Engle RF (2004) A Discrete-state continuous-time model of financial transactions prices: the ACM-ACD Model, forthcoming in J Bus Econo Stat
Tay AS, Ting C, Tse Y, Warachka M (2004) Transaction-data analysis of marked durations and their implication for market microstructure, Singapore Management University SESS Working Paper
Vergote O (2005) How to match trades and quotes for NYSE stocks? Center for economic studies discussion paper, K.U.Leuven, DPS 05.10
Wand MP, Jones MC (1995) Kernel smoothing, Monographs on statistics and applied probability 60, Chapman & Hall/CRC